

“Implementation on Health Care Database Mining in Outsourced Database”

Monali Gainkar¹, Prof. Sonali Bodkhe²

¹ Departement of computer science and engg,
G.H. Raisonni Academy of engg. & technology, Nagpur
monalimgainkar@gmail.com

² Departement of computer science and engg,
G.H. Raisonni Academy of engg. & technology, Nagpur
sonali.bodkhe@raisonni.net

Abstract: Due to the EMR (Electronic Medical Record) system there will be a rapid growth in health data collection. As we have already discuss in previous review paper the different work of the health care data record for maintaining the privacy and security of health care most private data. Now in this paper we are going to implement sheltered and secretive data management structure that addresses both the sheltered and secretive issues in the management or organization of medical data in outsourced databases. The proposed framework will assure the security of data by using semantically secure encryption schemes to keep data encrypted in outsourced databases. The framework also provides a differentially-private query or uncertainty interface that can support a number of SQL queries and complicated data mining responsibilities. We are using a multiparty algorithm for this purpose. So that all the purpose is to make a secure and private management system for medical data or record storage and accesses.

Keyword: Outsourced database, data mining, multi-party algorithm, data encryption.

I. INTRODUCTION

Healthcare industries store a enormous amount of perceptive personal data, such as patient names, dates of birth, and personal medical records. Since healthcare data doubling every year, organizations need to provide in both hardware and software to store and manage large amount of data. Database outsourcing has gained significance in the past few years due to the appearance of the cloud computing. In Database-as-a-Service (DaaS), which is a sort of cloud computing services, the database administrator outsources both databases and querying services to a cloud server and clients issue queries over the database to the cloud server. In this context, privacy is a most important test and it is necessary to satisfy main privacy requirements of database owners and clients. In the budding cloud computing archetype, data owners become progressively more aggravated to outsource their complex data

management systems from confined sites to the commercial public cloud for great elasticity and financial savings. For the contemplation of users' privacy, susceptible data have to be encrypted before outsourcing, which makes valuable data utilization a very tough task. In this domain, cloud computing is an effective solution for healthcare companies to handle huge amounts of medical records. However, healthcare organizations face two technical challenges. First, data outsourcing exposes sensitive healthcare data to un-trusted cloud service providers. Unauthorized access to sensitive medical records can have a significant negative impact on healthcare services. To ensure the confidentiality of the medical data stored on the cloud, we should depend on semantically-secure encryption schemes. Using semantically-secure encryption schemes, it must be infeasible for a computationally-bounded adversary to derive significant information about a message when given only the cipher-text and the corresponding public key. In this regard, the challenge is how to ensure data confidentiality while allowing query execution over encrypted data.

Second, driven by mutual benefits and regulations, there is a demand for healthcare organizations to share patient data with various parties for making inquiries purposes. Healthcare organization may allow data analysts (e.g., researchers) to execute aggregate queries and perform some data analysis tasks (e.g., classification analysis) on the database. In this regard, the challenge is how to support aggregate queries or complex data mining tasks on encrypted data while preventing inference attacks.

There have been a lot of research suggestions that independently address these two challenges. Most of the previous suggestions on secure outsourced databases suggest encrypting the data before moving it to the cloud. While encryption can provide data secrecy, it is of little use in deterring inference attacks. Similarly, there is a widespread literature on private data analysis. However, all these proposals require access to unencrypted data to generate privacy-preserving answers and therefore do not satisfy the data confidentiality requirement. This reality demands a new

privacy-enhancing technology that can simultaneously provide data confidentiality against an un-trusted database server, and prevent inference attacks from data analysts. We are proposing a general framework for secure and private data management in order to support effective data mining. The contributions of the approach are summarized as follows:

Based on real-life healthcare scenarios, we will first identify a new problem of secure and private data management of outsourced databases for data mining purposes

- We adopt a new privacy-enhancing protocol that can provide data confidentiality against an un-trusted cloud server by using semantically-secure encryption schemes. We then extend the protocol to support aggregate queries or complex data mining tasks on encrypted data while preventing inference attacks.
- Taking decision tree learning as an example, we will show that it is possible to compute a classifier on the encrypted data. The computed classifier provides differential privacy guarantee to prevent an inference attack.

II. LITERATURE REVIEW

By reviewing the previous work done by many people we are just taking into account their most precious work. One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold.

F. McSherry [1] has developed Data records that are protected from the data miner without enlightening original or secret record information or data. Lindell and Pinkas [2] has the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners and showed how to securely build an ID3 decision tree when the training set is distributed horizontally. S. Barouti, D. Alhadidi and M. Debbabi [3] has presents protocols for executing keyword search and combined SQL queries that maintain the confidentiality of both the client and the database owner. N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou [4] for the first time ever, define and solve the dilemma of privacy-preserving query over encrypted graph-structured data in cloud computing (PPGQ), and establish a set of authoritarian privacy necessities for such a secure cloud data deployment system to become a reality. F. Chen and A. X. Liu [5] has reflect on a two-tiered sensor network structural design in which cargo space nodes gather data from in close proximity sensors and answer queries from the descend of the network. M. A. ALzain and E. Pardede [6] have suggested the design of a new reproduction suitable for NetDB2 architecture, known as NetDB2 Multi-

Shares (NetDB2-MS). R. Mishra, D. P. Mishra [7] principally highlights some foremost security issues obtainable in current cloud computing surroundings. They aims at concurrently achieving data secrecy while still keeping the balancing relations intact in the cloud. His proposed system facilitates the data owner to assign most of computation intensive tasks to cloud servers without reveal data stuffing or user access right information.

III. PROPOSED METHOD

From the above helpful and most accurate discursion on previous work done by a great people about outsourcing the data in encrypted form in cloud for some most important benefits and studying and grabbing idea related to database query processing help to build our system.

Our Aim is to propose a sheltered and confidential data management structure that addresses. The proposed framework makes certain the security of data by using semantically-secure encryption system to keep data encrypted in outsourced databases. In this scenario the data in the form of datasets are stored in database act as a container and datasets are health care data which is very sensitive data that the data owner or organization of data doesn't want to reveal to next unauthorized party while outsourcing data in the cloud.

The framework also provides a differentially-private query interface that can maintain a number of SQL queries and difficult data mining tasks. In this scenario client can query to the server and apart from all this server can copied only the information or sensitive data to client only as per query held by him not more than that information is provided by the server. Because the data is very sensitive and this data kept in protection of the server side.

The Framework will address both the security and privacy issues in the management of medical data in outsourced database. In the face of many benefits, data collected works and allotment have become a big concern as it intimidate individual privacy. Our Idea is to propose a secure and private data management framework that addresses both the security and privacy issues in the management of medical data in outsourced database. The proposed framework ensures the security of data by using semantically-secure encryption schemes to keep data encrypted in outsourced databases. The framework also provides a differentially-private query interface that can support a number of SQL queries and complex data mining tasks.

For Private access of Outsource data we are proposing a protocol which is based on two secure multi-party Algorithm:

- One for computing the union (or intersection) of private subsets that each of the interacting players hold:
 In this we have a dataset of health care and we perform classification our datasets by making union of datasets or interaction of data sets and as a result of it the dataset we get is a common datasets and un-common datasets which is hold by a party or user which gone access this database.
- Another is a protocol that tests the inclusion of an element held by one player in a subset held by another:
 After first step we are applying inclusion algorithm that can make a between the datasets held by one party with other one.

Secure Multi-party algorithm:

- A set of parties with private inputs wish to compute some joint function of their inputs.
- Parties wish to preserve some securityv properties. E.g., privacy and correctness.
- Security must be preserved in the face ofv adversarial behaviour by some of the participants, or by an external party.

Algorithm:

Range of the inputs known: (0, N)

Recap:v

A : public key e, private key dv

B : can access e, but not dv

And

- $Dec_d(Enc_e(x)) = x$
- $Dec_d(Enc_e(x)+y) = \text{random looking thing}$ (for someone who knows x,y,e but not d)

Step 1:

- A has i and B has j
- B generates a random x (of m bits)
- $C = Enc_e(x)$
- $u = C - (j-1)$
- Send u to A.

Step 2:

- A computes: for (t= 1 to N)
 $y_m = Dec_d(u+t)$
- Takes a prime p (of size about \sqrt{m}) and computes
 $z_i = y_i \bmod p \dots$ for i = 1 to N
- p chosen such that $|z_m - z_n| \geq 2$ for any m,n in [1 to N]

Step3:

- A send to B the following list: P, $z_1, z_2, \dots, z_i, (z_{i+1}+1), (z_{i+2}+1), \dots, (z_{N+1})$.
- Bob compares the j^{th} entry of this list (excluding the prime p) with x mod p.
- If $x \bmod p$ is j^{th} entry of the list implies $i \geq j$.

IV. EXPERIMENTAL RESULT

There will be a three important person in this project oh health care data mining in outsources database.

1. Client
2. Owner
3. Server

OWNER Login:

D8S1179	D21S11	D7S820	CSF1PQ	D3S135	D13S31	C16S56	D2S133	D19S43	D5S812	ClientID	TestOfMonth	DataVerify
0.7	0.4	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.23	jon	Jan	NO
0.7	0.4	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.23	123	Jan	NO
0.2	0.5	0.3	0.6	0.7	0.6	0.4	0.6	0.8	0.5	kitty	May	YES
0.3	0.025	0.36	0.21	0.36	0.24	0.29	0.15	0.36	0.31	petter	Jan	NO
0.24	0.23	0.09	0.26	0.28	0.24	0.27	0.26	0.24	0.35	dora	Feb	NO
0.24	0.23	0.09	0.26	0.28	0.24	0.27	0.26	0.24	0.35	dora	Feb	NO
0.14	0.25	0.36	0.28	0.39	0.45	0.26	0.27	0.28	0.29	kittu	Apr	NO
0.14	0.25	0.36	0.28	0.39	0.45	0.26	0.27	0.28	0.29	kittu	Apr	NO
0.25	0.14	0.25	0.36	0.17	0.08	0.09	0.04	0.05	0.36	pinky	Jun	NO

When client registered his data and submit it the data owner get all the client data as we can see in the screen.

DNA Locus	Results	DNA Locus	Results
D8S1179	>>>> [input]	D13S31	>>>> [input]
D21S11	>>>> [input]	C16S56	>>>> [input]
D7S820	>>>> [input]	D2S133	>>>> [input]
CSF1PQ	>>>> [input]	D19S43	>>>> [input]
D3S135	>>>> [input]	D5S812	>>>> [input]

Then data owner have to enter data for each and every client.

SERVER Login:



Health Care Database Mining in Outsourced Database

Server Home Data Key Verify Data Association Rule View Computation Log Out

SERVER HOME PAGE

Welcome SERVER!

View CLIENT Encrypted DNA Test Result's Here !

DNA ID	D8S1179	D21S11	ClientID	TestOnMonth	DataVerify
1	K9nsvGqLjv6zgrI3AMgiaw==	qsdi7bUwWgnLVXog8pDNA==	jon	Jan	NO
2	K9nsvGqLjv6zgrI3AMgiaw==	qsdi7bUwWgnLVXog8pDNA==	123	Jan	NO
3	0HopsQNpNYIXyBhS6YRg==	Pa2Rh2YK7m/RmWQgZUA==	katy	May	YES
4	HcxwKKKpHwJ-WFL-57EA==	xCB9PacoSKdeg qpEv-g==	getter	Jan	NO
5	4e1Hci3chBi-0CpTTB4W3A==	KH4gDzuru1U7Eh6UMn91A==	dora	Feb	NO

1 2 3 4 5 6 7 8 9 10 ...

When server have logged in into system he get all the data entered by data owner by registered client in encrypted form.



Health Care Database Mining in Outsourced Database

Server Home Data Key Verify Data Association Rule View Computation Log Out

DNA DATA KEY

Welcome SERVER!

View Data Key Here !

Select any Data ID: 11

Encrypted Data Key is : XobXrjFAPra4d24ZTNr4FWw==

Get Decrypt Key

Decrypted Data Key is : 32184

And encrypted data of each client DNA key is get by ID to verify each client data for security purpose. The entire association rule mining task is done by server side.

And there will be three type of computation result will shown for data owner, DNA entries, data verify in graphical format at the end.

Computation Result for data owner:



Computation Result for DNA entries:



Computation Result for data Verification:



V. CONCLUSIONS

Our main goal is to make a system which provides a security and privacy preserving task in outsource database. An making query to database server and only that much information is provided to user. The proposed framework will ensures the security of data by using semantically-secure encryption schemes to keep data

encrypted in outsourced databases of the health care database. This is more sensitive data for organization. The framework will also provide a differentially-private query interface that can support a number of SQL queries and complex data mining tasks. Our Multiparty Protocol will also guarantee the private access of outsourced data.

VI. REFERENCES

- [1] F. McSherry, "Privacy integrated queries," in Proceedings of the 35th ACM International Conference on Management of Data (SIGMOD), 2009.
- [2] Y. Lindell and B. Pinkas, "Privacy preserving data mining," *Journal of Cryptology*, vol. 15, no. 3, pp. 177–206, 2002.
- [3] S. Barouti, D. Alhadidi, and M. Debbabi, "Symmetricallyprivate database search in cloud computing," in *Cloud Computing Technology and Science (CloudCom)*, International Conference on, vol. 1. IEEE, 2013, pp. 671–678.
- [4] N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou, "Privacy-Preserving Query over Encrypted Graph-Structured Data in Cloud Computing," 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, 2011, pp. 393-402.
- [5] F. Chen and A. X. Liu, "SafeQ: Secure and Efficient Query Processing in Sensor Networks," 2010 Proceedings IEEE INFOCOM, San Diego, CA, 2010, pp. 1-9.
- [6] M. A. ALzain and E. Pardede, "Using Multi Shares for Ensuring Privacy in Database-as-a-Service," 2011 44th Hawaii International Conference on System Sciences, Kauai, HI, 2011, pp. 1-9.
- [7] R. Mishra, D. P. Mishra, A. Tripathy and S. K. Dash, "A privacy preserving repository for securing data across the cloud," 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari, 2011, pp. 6-10.
- [8] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacypreserving data publishing: A survey of recent developments," *ACM Computing Surveys*, vol. 42, no. 4, pp. 1–53, June 2014.
- [9] C. Dwork, "Differential privacy," in *Proceedings of the International conference on Automata, Languages and Programming (ICALP)*, 2012.
- [10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the 3rd conference on Theory of Cryptography (TCC)*, 2008.
- [11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann Publishers Inc., 2011.
- [12] M. D. Berg, O. Cheong, M. V. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, 3rd ed. Springer-Verlag TELOS, 2008.
- [13] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [14] Z. Zhu and W. Du, "Understanding privacy risk of publishing decision trees," in *Proceedings of the 24th Annual IFIP WG*
- [15] *Working Conference on Data and Applications Security and Privacy (DBSec)*, 2010.
- [16] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, "A practical differentially private random decision tree classifier," *Trans. Data Privacy*, vol. 5, no. 1, pp. 273–295, Apr. 2012.